Scholarly Publishing and Argument in Hyperspace

Victoria Uren
v.s.uren@open.ac.uk

Simon Buckingham Shum
s.buckingham.shum@open.ac.uk

Gangmin Li
g.li@open.ac.uk

John Domingue
j.b.domingue@open.ac.uk

Enrico Motta
e.motta@open.ac.uk

Knowledge Media Institute
Open University
Milton Keynes, MK7 6AA, UK

ABSTRACT

The World Wide Web is opening up access to documents and data for scholars. However it has not yet impacted on one of the primary activities in research: assessing new findings in the light of current knowledge and debating it with colleagues. The ClaiMaker system uses a directed graph model with similarities to hypertext, in which new ideas are published as nodes, which other contributors can build on or challenge in a variety of ways by linking to them. Nodes and links have semantic structure to facilitate the provision of specialist services for interrogating and visualizing the emerging network. By way of example, this paper is grounded in a ClaiMaker model to illustrate how new claims can be described in this structured way.

Categories and Subject Descriptors

H.5.4 [**Information Systems**]: Information Interfaces and Presentation - *Hypertext/Hypermedia;* H.3.7 [**Information Systems**] Information Storage and Retrieval - *Digital Libraries*

General Terms

Design, Human Factors.

Keywords

Scientific Publishing, Scholarly Interpretation, Collaborative Web, Modeling Debate.

1. INTRODUCTION

Debate at a distance is integral to modern scholarship. Many researchers participate actively on list servers for their specialism, and more "lurk" on the lists of related subjects to keep abreast of the issues that are exciting people there. At the Knowledge Media Institute, with the Open University's distance learning students in mind, we take a long term interest in technologies that can help support debate at a distance. We have developed a variety of tools for supporting scholarly interaction. These include D3E [2; 25], a system for the discussion of single documents which has been implemented in the open peer reviewed journal JIME [13].

For most scholars the World Wide Web is a document repository, plus a convenient way to access datasets and services. As such it is a valuable resource and services like CiteSeer [5] and Google [3] are constantly pushing the boundaries of what can be done with information retrieval technologies. The Semantic Web promises a future in which intelligent services will be provided through the markup of Web documents according to agreed conventions [10]. The question we address is whether the current Web as document repository and the emerging Semantic Web can be brought together to create something more like the scholarly hypertext envisaged by Kolb [21]. This would be an environment in which scholars could debate and comment on articles, in the process building a model of the literature under analysis. The field of argument visualization, which is attracting interest in educational technology, public policy and organizational computing, is relevant to this challenge [14].

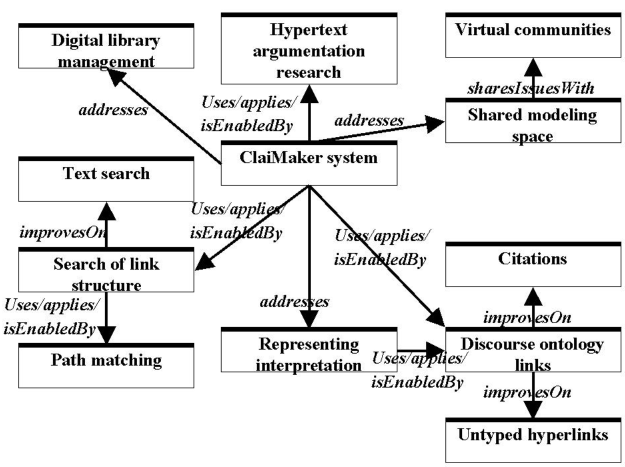

Figure 1 Schematic diagram of a ClaiMaker model for this paper "Scholarly Publishing and Argument in Hyperspace". Boxes represent concepts and arrows represent typed links. A Concept-link-Concept triple is a "claim".

In the Scholarly Ontologies Project we are building a prototype tool called ClaiMaker [24] that can be used to summarize the core contributions of an article, and their links to other work. This builds on the World Wide Web as document repository by providing URLs that link parts of the model to the document of interest. Furthermore, the models use typed links which may be exploited in the provision of intelligent services.

This document has a slightly unusual format because we have "taken our own medicine". As a demonstration that the core of an article can be represented in the form of a ClaiMaker model we started with the model (Figure 1) and built this paper from it. Subsections deal with parts of the model, and are clustered in sections based on their proximity. There is no particular reason to

WWW 2003, May 20-24, 2003, Budapest, Hungary.
ACM 1-58113-680-3/03/0005.

read the document linearly, but if you choose to it should still be comprehensible.

1.1 ClaiMaker system *uses/applies/ isEnabledBy* Hypertext Argumentation Research

It is widely recognized that hypertext was shaped by the Memex vision of Bush [16] and the NLS system of Engelbart [18]. It is less commonly known that both of these pioneers saw the construction and analysis of scholarly arguments as key applications of their technologies. The work described here can thus be traced back to Bush and Engelbart, via the extensive research in the 1980's and early 1990's into hypertext graphical argumentation tools, and more recent work on hypertext scholarship (for a review see [14]). ClaiMaker models are analogous to signposts on the route of Bush's associative trails where the terrain being crossed is the Internet. This progression from the closed pre-Web argumentation systems to the Internet increases the scale of the user community proportionally.

2. CLAIMAKER SYSTEM

ClaiMaker is a prototype tool for modeling documents, that combines knowledge engineering and visualization. It is being developed and assessed as part of the Scholarly Ontologies project. The current version is a server side application. This facilitates getting new versions to early adopters and also gives us access to the models they build allowing us to assess the modeling scheme and identify difficulties.

The core functionalities of ClaiMaker are model building and discovery. The model building activity is document centric. Each concept (a box in Figure 1), and each triple of two concepts joined by a typed link (an arrow in Figure 1), is attributed to a document. This provides backing. In the example being discussed here the components of the model are backed by this document. Users are encouraged to make links to concepts backed by other documents and to reuse concepts. They may extend the models built by other contributors, adding further claims, or take issue with them if they feel the original interpretation is flawed. Thus the claim space emerges collaboratively and cumulatively as a complex web of interrelated claims.

Claims are stored in a mySQL database which supports all the basic functions of model building, and search. The database is accessed via a webserver. A mirror of the database in RDF is also maintained on a development server where we are designing services that exploit the structure in the link ontology (see section 4). The first of these RDF-based services will soon be publicly available.

Discovery comprises searching the claims network and presenting the results to users. Developing discovery services is a core research and development activity within the project. One of the approaches taken will be discussed in Section 6. Work on the visualisation of claim structures is presented elsewhere [15].

3. DIGITAL LIBRARY MANAGEMENT

The Web as document repository has a heterogenous structure and dispersed content. Sources such as digital libraries and pre-print archives apply some standardization to texts and data collections as well as widening access by removing the physical and geographical constraints on librarians and readers. However, bringing distributed archives to distributed users requires a considerable effort in managing resources and making them available. Standardization activities such as Z39.50 [9] and the Open Archives Initiative [6] are dealing with core issues of infrastructure and interoperability and diverse work is being done by e-librarians and software developers to make user services. Nonetheless solitary users working like detectives to extract the information they need from multiple sources face usability difficulties [11].

3.1 ClaiMaker system *addresses* Digital Library Management

ClaiMaker allows groups of researchers to collate their discoveries from digital libraries in a meaningful way. It gives control over the representation of the document collection to the users rather than to the managers of the collections. The users model their interpretations of the documents, producing a picture of the documents that is tailored to that particular group.

Contributions to the model are "backed" by documents. These could be held in a specific digital library, or they could be links to distributed sources. They could even be conventional citations of non-digital sources.

As a group builds a model they effectively build an index aimed at their peers and grounded in their readings of the literature. Group members can benefit from the collective knowledge of the group; they are less solitary. Clearly the dynamics of groups that would make knowledge available to each other are going to be rather special. These issues are discussed further in section 5.

4. REPRESENTING INTERPRETATION

Scholarly documents, even those that present factual results from experiments, are not bald collections of data but arguments. The authors have to make a case that the work they are presenting is a significant contribution and that their conclusions are reliable, given the current state of knowledge. The scholarly article can be seen as an *interpretation* of data in the light of wider debate.

4.1 Representing interpretation *uses/applies /isEnabledBy* Discourse Ontology Links

Representing interpretation is a knowledge engineering problem. By the nature of research we expect that the topics in which people are interested will shift as they take up new challenges. Additionally, the kinds of conceptual objects which are core to the research process will differ from discipline to discipline. The application of ontologies to contested fields may not seem a promising strategy when ontologies typically depend on consensus to control interpretation. However, the mechanisms of scholarly debate do remain stable over time, making *discourse/argumentation* ontologies good candidates for representational support. Whether research is in the arts or sciences there will always be *problems* that are of key interest, people will put forward *theories, predictions, hypotheses,* etc., and try to support them with *data* and analysis. These contributions may, in their turn, be challenged or developed further. Therefore our knowledge modeling effort has focused on capturing enduring, discipline-independent *relationships* between objects, rather than the *types* of objects. We have developed an ontology of links to represent the rhetorical moves researchers make when they present their arguments. Domain-specific vocabularies (e.g., the ACM Computing classification Scheme) could be added if desired, but are orthogonal to the discourse ontology.

This discourse ontology is a structured set of suitable links that provides a common language for authors to use when making claims and positioning them with respect to earlier work. It differs from most ontologies because it must accommodate more than one view of the data, because researchers often have different or even conflicting perspectives. The ontology provides the structures that are being used to design innovative search, interpretation and visualization tools to allow users to navigate large and growing models.

The discourse ontology supplies a palette of typed links. The links are used to join objects we call "concepts", which are short pieces of text describing a problem, a piece of experimental data, a hypothesis etc. The resulting triple we call a "claim"; it is in effect the assertion that a particular relationship holds between two concepts. To imagine how such claim triples might be claimed to be "interpretations" consider the following self-criticism of the triple used as the header for this subsection. *Representing interpretation* is a problem the Scholarly Ontologies Project is trying to tackle. We have developed a set of Discourse Ontology Links. Do these links really *enable* the representation of interpretation? We claim they do, because if you the reader write a paper proposing a better way of representing interpretation you can link your paper into this one *challenging* our claim, or *extending* our work. Your interpretation and ours would then coexist in the same claim space.

Claims were modeled on paper for a range of research domains, including computer supported collaborative work, text categorization, and literary criticism. Relations common to several domains were identified. We found we could classify these into groups with similar rhetorical implications: Supports/Challenges, Problem-related, Taxonomic, Causality, Similarity, and General. Each relation belongs to one group. We also found that some relations occurred in pairs of opposites, e.g., *proves* and *refutes*, where one has positive and the other negative implications. We call this property "polarity". For example, *refutes* has negative polarity; it implies *dis*proof. Some relations are more forceful than others, for example *refutes* has stronger implications than *disagreesWith*. Therefore each relation is also assigned the parameter "weight".

Table 1 gives a breakdown of link characteristics for the Supports/Challenges class. It can be seen that the links occur in pairs of positive and negative links with opposite implications. Furthermore, two of the links, proves and refutes, are of high weight. If an author uses these the reader would expect strong evidence. To illustrate discourse ontology links in use, we will examine some of the triples from the model on which this paper is based.

[ClaiMaker system] (*addresses*) [Digital library management]

This uses one of the Problem-Related relations *addresses,* to assert that *ClaiMaker* is looking at an established problem, *Digital library management*. If the model were part of a real claim space the latter concept could have already been introduced by a different document, reusing the concept would position the model for this paper in context of other Digital Libraries research.

[Shared modeling space] (*sharesIssuesWith*) [Virtual communities]

This claim uses a relation of type Similarity to link into an established field. The implications are different. Here we say that some of the social issues of running a shared modeling space with a tool such as ClaiMaker have already been examined in a different context. In subsection 5.1 we talk about specific issues in more detail.

[Discourse ontology links] (*improvesOn*) [Untyped hyperlinks]

This claim uses a General relation. It has the parameters strong and positive (see subsection 4.2 for further discussion of this claim).

The Discourse Ontology is intended for use by subject specialists, not by trained knowledge engineers. Therefore, within the ClaiMaker tool the links are selected by the modeler as labels, and the polarity and weight of links are "hidden" so as not to burden the user with having to understand and assign weights. The parameters come into their own in the Discovery stages where they are used as the basis of services. They also mean that the tool can be modified for research communities that use a different vocabulary, or a different language, to conduct debate. The labels in the new "dialect" may be changed, but services which operate on type, polarity and weight parameters will not be affected.

Table 1. Table summarizing discourse ontology parameters for the Supports/Challenges class of links

Label	Polarity	Weight
proves	positive	high
refutes	negative	high
isEvidenceFor	positive	low
isEvidenceAgainst	negative	low
agreesWith	positive	low
disagreesWith	negative	low
isConsistentWith	positive	low
isInconsistentWith	negative	low

When developing the discourse ontology, we were aware of a trade-off between usability and expressiveness. Users may be wary of a very complex system, preferring not to use it rather than be seen to make a "mistake". On the other hand very simple systems impose too many constraints on the types of models users can build. We took a pragmatic approach to selecting relations, aiming for a moderate palette of useful links rather than a very complete set such as that proposed by Trigg [27]. Since then Gil and Ratnakar have published work on the TRELLIS system [19] which also uses discourse relations. We note that they too selected relations that could be understood by users, rather than "precise" or "complete" relations. Our ontology differs from theirs in that we are also striving for links that are computable in order to help manage and interrogate the emerging structures.

In addition, we argue that having a relatively simple ontology makes the system more robust to imperfect modeling. Since the proposed user group is researchers, who are adept at structured thinking but are not expert knowledge modelers, we expect a degree of imperfection. If, for example, a user chooses to use the *agreesWith* link where the *isConsistentWith* link would be more appropriate the polarity and type of the link remain the same and so services that use those parameters are robust to the error. Moreover, even if completely the "wrong" relation is used, the fact that someone has made a connection of some sort will be used in services that ignore semantics.

4.2 Discourse Ontology Links *improvesOn* Untyped Hyperlinks

The Web has shown conclusively that linking documents in a non-linear way can enhance information. It has also demonstrated that an anarchic information space in which anyone can add any document and link it to any other faces major usability challenges. As pointed out by numerous hypertext researchers, part of the problem is that on the Web all hyperlinks are alike. This hampers human navigation to some extent, although the current generation of Web pages are written to be interpreted by people who can guess where a link may lead from its context. The main problem of untyped hyperlinks is rather that they limit the types of automatic processing that can be done.

The Semantic Web [10] is pertinent here. The aim of Semantic Web activity is to make the content of web pages machine interpretable. This is to be done with information structured using markup languages, and schemas and ontologies written in RDF, OIL etc. Since URLs can be used within these structures they effectively supply a mechanism for making typed hyperlinks. This opens up possibilities for reasoning and semantic services.

From a Semantic Web perspective, the Scholarly Ontologies Project can be recast as the development of an annotation schema and associated services. Annotation consists of creating machine interpretable structures that reference Web documents [7]. With this annotation model in mind we can envisage a scenario in which ClaiMaker models are published alongside the online versions of research papers. These could be downloaded by readers and meshed into their personal or community repositories, possibly further enhanced with personal comments. Software agents could be dispatched to crawl the Web to find models that match a specific query, pooling the results and analyzing them.

The key to proving this claim is to demonstrate that typed links can be used to design semantic services. We will discuss this further in section 6.

4.3 Discourse Ontology Links *improvesOn* Untyped Citations

In ClaiMaker models claims have to be backed by linking to a source document that discusses the issues in depth. In research papers, backing may be provided by citing other work. These models of backing share issues with the analysis of argument structures presented by Toulmin [26].

Citations are proven research tools. Users of CiteSeer will know the benefits of being able to navigate from a paper to the articles it cites and the articles it is cited by [5]. The development of such linking services within the Open Archives framework was explored by the OpCit Project [20]. The Science Citation Index as established in the 1950s both as a tool for literature searching and as a repository of data that could be analyzed to get high level insights into the progress of science [4]. Citation data continues to fuel useful research, for example to reveal the connections between researchers in a domain [17].

However, as with hyperlinks, all the above examples use untyped citations. In practice, the rhetorical use of citations is a subcraft of research writing. The citation in the first paragraph of this section has an explanatory purpose. “Backing” is not an obscure word, but by providing the link to Toulmin’s notion of backing readers who already know his work can rapidly orient themselves, for others it provides a signpost to a source of further data. Weinstock identified paying homage, correction and substantiating claims amongst others in his classic taxonomy of motivations for citing [28]. A limitation of traditional citation studies is that they do not leverage such nuances of motivation, indeed they cannot even distinguish between citations that support the cited work and those that oppose it.

By contrast, the links from concepts backed by one paper in ClaiMaker to concepts backed by another have both *labels* (i.e., they are typed citations), from which the motivation can be judged by humans, and computable features such as *polarity* and *weight*. As an example of how polarity can be put to work, ClaiMaker contains a discovery tool called Contrast which can identify concepts that have negative links proximate to concepts backed by a particular paper [24]. This is implemented as a sequence of three SQL searches:

1. Select concepts backed by the paper
2. Extend the set by selecting concepts that are linked to/from the first set by a positive link
3. Select all the claim triples that include one of the concepts in the extended set and a link with negative polarity

Although it is simple, the Contrast service demonstrates that link parameters can be used to provide services to users.

5. SHARED MODELING SPACE

The ClaiMaker tool is intended to serve a community of domain experts, who may be the authors of the documents being modeled or readers. Each presents their interpretation of claims made by relevant documents and links them to related models. However the tool itself can only provide mechanisms for handling data. The model is the common property of the community and the tool is only of value if it fits a need and works within existing mores.

5.1 Shared modeling space *sharesIssueswith* Virtual communities

Researchers belong to communities in which documents play a central role. They are very often introduced to each other through their published work long before they meet in person. Schools of thought frequently coalesce around seminal papers. A typical example of such a document is Bush’s Memex article [16]. Research domains differ in their methods and ambitions but they all engage in constant reading and production of documents, as communication tools and as a lasting open record of their results. Tools for virtual research communities should support document centric debate at a distance.

It has been noted that science communities have been virtual communities from their inception [12], and the same would apply to research communities in general. Although much of the research about virtual communities has centered on recreational systems such as MUDs and graphical virtual worlds, less glamorous technologies are providing foci for scientific communities. For example, the September98-Forum list server [8] provides an example where scholars, are debating, often passionately, the role of free to read electronic archives in scholarly publishing. However, the very success of some list servers can limit their usefulness. At the time of writing the

September98-Forum list had 225 streams of discussion for the year 2002, the longest stream had 31 email messages, and approximately 40% of the messages were over 100 lines long (the longest was 1164 lines). Following this debate is a formidable task, coming to it for the first time is daunting. Virtual research communities require a medium that will support them in producing summaries of their debates. We believe that ClaiMaker is a prototype of a summarizing medium.

Identity persistence is necessary for successful cooperative relations [22]. In the context of ClaiMaker, identity means that users cannot add to the debate anonymously; all concepts and claims are clearly labeled with the details of the creator as well as the original authors of the paper. When creating a concept node to refer to someone else's work, the user also indicates whose 'intellectual property' they are claiming it is: are they merely intending to make explicit an idea already in the paper (IP=DocAuthors), or are they claiming to read something new into the paper? (IP=ClaiMakerUserID). Identity lets information users make judgements about trust. Whose claims about a paper do you trust most: the author's, an unknown contributor's, or a world expert's? It also brings accountability: what claims are you willing to make in the public domain? Accountability provides the foundation for informal social controls on what is acceptable in the community.

The issue of social controls leads us to ask what kind of virtual communities might use ClaiMaker? What degree of control might be placed on contributors? How public will they wish to be? We currently have several claim spaces in operation. The most public is the Sandpit [1]. Anyone can browse the models in the Sandpit or create an account to establish their identity, and start making their own models. However, we do not propose this claim space as a suitable area for serious modelers. It is more a place for people to familiarize themselves with the tool and decide if it is for them. Committed users may require a more private space. Whom they invite to join them there will then depend on the structure of their community in the wider world. A community of researchers in a particular domain may open the claim space to known colleagues. A research group may build a claim space of key papers in the area, but keep the resource private if they consider it gives them a competitive advantage. Lecturers may require students to contribute to a model of the papers on the course reading list as an assessment exercise.

The ultimate question for a new technology is "will people use it?" This applies particularly to systems like ClaiMaker where busy users would have to expend time and effort on a new task. We can learn here from the cost benefit trade-off offered by the Web itself. The Web has been likened to a gift economy. Scholars put lists of their publications on their home pages, appreciating the ready access they have to other people's lists, and hoping that they will be read and even cited. If scholars are to publish ClaiMaker models in a communal space they will have to feel that they will get some benefit. One form would be the social benefit of increased digital visibility and recognition by other members of the community. This is not a dynamic that can be manipulated technologically any further than by ensuring contributors get credit for their work, which is an identity issue. Further benefits are access to other people's models and interpretations, the ability to distill personal views from a larger model, and agent mediated alerting to potentially relevant new work. We are still in the early phases of this kind of scholarly publishing, making it hard at this stage to deliver the critical mass that is required for people to adopt new web services. Modeling is done by researchers and early adopters who see its potential. In the longer term, maximizing the cost-benefit tradeoff is the key challenge. It is by providing task-oriented tools that help researchers add to and query the claim space that we hope to demonstrate that the value of contributing to a communal model.

6. SEARCH OF LINK STRUCTURE

A ClaiMaker model comprises a collection of short pieces of text (concepts) joined using links whose characteristics are defined by the discourse ontology described in section 4.1. It is the link structure, more than the text in the concepts, that drives the design of discovery tools.

6.1 Search of Link Structure *uses/applies/isEnabledBy* Path Matching

The enabling technology for searching link structure is path matching. This is a graph theoretic method which searches for paths (sequences) of links that follow a specific pattern. To experiment with this approach we expressed a ClaiMaker model of about 700 concepts and 600 links in RDF and used the Ivanhoe path matcher embedded in the Wilbur RDF Parser [23].

6.2 Search of Link Structure *improvesOn* Text Search

Searching text is a very well established information retrieval technique which has been refined over decades and is known and understood by millions of users worldwide. Setting ourselves up to improve on text search could be seen as hubris. Seen from another perspective, text still poses hard problems. For example, in the diverse environment of the Web, a word that is used in different contexts can return many hits, most of which are irrelevant. From a Scholarly Ontologies perspective, text search engines have a deeper flaw: they deal only with the words in documents, they do not support interpretation of documents. This limits the kinds of questions that can be posed. For example, it is possible to ask for documents on a particular topic, but not to ask how the topic evolved. We have taken on the task of answering questions such as "Has anyone built on this idea?" and "Are there distinct schools of thought on this issue?" that require interpretation. Text search cannot address this kind of query.

We illustrate our present approach to answering questions that require interpretation by looking at the question "where did this come from?" This relates to a common activity in research, clarifying the lineage behind an idea. Lineage, and its inverse, the descendant, focus on the notion that ideas build on each other. We set the goal of picking out from the "spaghetti" of claims, candidate streams of ideas that appear to be building on each other. Our lineage tool tracks back semantically from a concept to see how it evolved, whereas the descendants tool tracks forward. Since descendants are the inverse of lineage, we will only discuss lineage.

Lineage was defined as a path in which the links suggest development or improvement. Path queries were constructed from link-types using a set of primitives. For example, we can search for paths that may be of any length, and which contain (in any order) any of the positive links that have type Similarity in either direction, or the two general links *uses/applies/isEnabledBy* or *improvesOn*, going in the direction away from the target concept of the query. The *improvesOn* link type is included to reflect the notion of progress implied by lineage, while *uses/applies /isEnabledBy* has a weaker implication of “building upon.” The Similarity links are included because if a new concept is like another that *improvesOn* a third concept, then the new concept is reasonably likely to be an improvement as well. Similarity links are acceptable in either direction because natural similarity is a symmetrical relation: if A is like B, then B is like A.

In Figure 2 the principle of this lineage search is illustrated using the model for this paper. The figure shows the lineage that would be extracted for the concept “*ClaiMaker System*”.

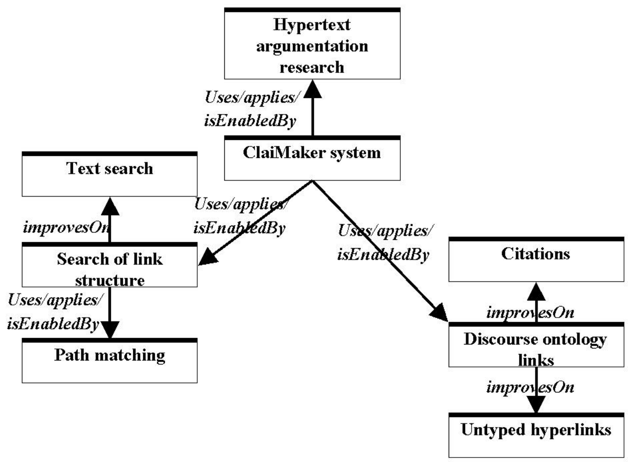

Figure 2 Schematic diagram of the lineage of the concept “ClaiMaker system”. Boxes represent concepts and arrows represent typed links. A Concept-link-Concept triple is a "claim".

The search can be tightened by filtering the paths returned to ensure they contain the *improvesOn* relation. A search could also be broadened, to permit the inclusion of any Problem-related links (since *addressing* or *solving* a known problem usually represents progress of some sort), or to include Taxonomic links (since if a *partOf* some innovation *improvesOn* another approach then it implies there may be an improvement overall). Note that in these cases, the direction of the link is fundamental: it is only problems that the new concept *solves* that are of interest, and even if a whole innovation is an improvement, there is no reason to assume that every *partOf* it is also.

7. CONCLUSIONS AND FUTURE WORK

Despite being central to all scholarly disciplines, interpretation and debate remain the province of verbal and textual communication. Meanwhile, the consumers of new research results must manage increasing amounts of information in order to stay up to date. There is a clear need for powerful, usable tools to prevent researchers from ‘drowning in the information ocean’. Consequently, the focus of our work is the evolution of the Web to better support scholarly publishing and interpretation.

We have described the Scholarly Ontologies project whose ClaiMaker system offers a network model for summarizing research debates over a whole literature as well as for individual documents. This can be done asynchronously by groups of distributed users who build their models on a central server, thus supporting debate at a distance. Furthermore, the models are machine interpretable, allowing us to develop novel user services such as analyzing the lineage of ideas.

At this point, we have an ontology of links for discourse and a system that supports model building and some core services. The future promises some interesting challenges. Increasingly, services will incorporate visualization to assist the human interpretation of models. We will continue to collect evidence to support the ClaiMaker approach. Seeking and fostering early adopter communities to use and help in the evaluation of the system is now a central activity. Closely associated with this will be the implementation of further services to give those communities payback for their contributions to models. The reverse side of that particular coin is reducing the cost of building ClaiMaker models. In the early days of the World Wide Web people built their HTML competence by cutting and pasting useful bits of HTML code from other people's web pages. For ClaiMaker we are proposing a series of extensible templates for particular genres of paper. Users can then start out by “filling in the slots” on a standard model before progressing to build their own models from scratch.

A subsidiary aim of this article was to demonstrate that the core of a research paper can be represented in the form of a ClaiMaker model. This has been a stimulating exercise. The model made the skeleton of the arguments clear. Putting flesh on the bones required explaining our meaning for concepts, and justifying our position for claim triples.

Finally, if this article has aroused your interest and you want to try using ClaiMaker, you are very welcome to come and play in the Sandpit [1].

8. REFERENCES

[1] ClaiMaker Sandpit. http://claimaker.open.ac.uk/Sandpit/

[2] Digital Document Discourse Environment. http://d3e.sourceforge.net

[3] Google. http://www.google.com/

[4] History of citation indexing. http://www.isinet.com/isi/hot/essays/citationindexing/21.html

[5] NEC Research Institute CiteSeer. http://citeseer.nj.nec.com/

[6] Open Archives Initiative. http://www.openarchives.org/

[7] RDF Annotations. http://ilrt.org/discovery/2001/04/annotations/

[8] September-98 Forum Archives. http://amsci-forum.amsci.org/archives/september98-forum.html

[9] Z39.50 International Standard Maintenance Agency. http://www.loc.gov/z3950/agency/

[10] Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. Scientific American (2001). 34-43.

[11] Blandford, A., Stelmaszewska, H., Bryan-Kinns, N. Use of multiple digital libraries: a case study. 1st ACM/IEEE-CS Joint Conference on Digital Libraries, Roanoke VA (2001).

[12] Brown, J. S., Duguid, P. The social life of information, Harvard Business School Press (2000).

[13] Buckingham Shum, S., Sumner, T. JIME: an interactive Journal for Interactive Media. Learned Publishing: Journal of

Association of Learned and Professional Society Publishers (2002). 14(4) 273-285.

[14] Buckingham Shum, S. The roots of computer supported argument visualization. Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. P. A. Kirschner, Buckingham Shum, S. and Carr, C. London, Springer-Verlag. (2003) 3-24.

[15] Buckingham Shum, S., Uren, V., Li, G., Domingue, J., Motta, E. Visualising internetworked argumentation. Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. P. A. Kirschner, Buckingham Shum, S. and Carr, C. London, Springer-Verlag. (2003) 185-204.

[16] Bush, V. As we may think. The Atlantic Monthly (1945). 176 101-108.

[17] Chen, C., Paul, R.J. Visualising a knowledge domain's intellectual structure. Computer (2001). 34(3) 65-71.

[18] Engelbart, D. C. Augmenting human intellect: A conceptual framework. SRI Project No. 3578, Summary Report AFOSR-3233 Stanford Research Institute.

[19] Gil, Y., and Ratnakar V. TRELLIS: an interactive tool for capturing information analysis and Decision making. EKAW 2002, LNAI 2473 (2002).

[20] Hitchcock, S., Carr, L., Jiao, Z., Bergmark, D., Hall, W., Lagoze, C., Harnad, S. Developing services for open eprint archives: globalisation, integration and the impact of links. ACM DL00, Fifth ACM International Conference on Digital Libraries (2000).

[21] Kolb, D. Scholarly Hypertext: self-represented complexity. Hypertext '97 (1997).

[22] Kollock, P. Design principles for online communities. PC Update (1998). 15(5).

[23] Lassila, O. Enabling Semantic Web programming by integrating RDF and Common Lisp. SWWS, Semantic Web Working Symposium, Stanford (2001).

[24] Li, G., Uren, V., Motta, E., Buckingham Shum, S., Domingue, J. ClaiMaker: weaving a semantic web of research papers. ISWC2002, 1st International Semantic Web Conference, Sardimia (2002).

[25] Sumner, T., Buckingham Shum, S. From Documents to Discourse: Shifting Conceptions of Scholarly Publishing. CHI98: Human Factors in Computing Systems, Los Angeles, ACM Press (1998).

[26] Toulmin, S. The uses of argument, Cambridge University Press (1958).

[27] Trigg, R. A network-based approach to text handling for the online scientific community. (1983) Department of Computer Science, University of Maryland

[28] Weinstock, M. Citation Indexes. Encyclopedia of Library and Information Science. (1971) 5, 16-40.